

# Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means

## (Clustering Sales Data at Outdoor Equipment Stores Using K-Means Method)

Fintri Indriyani<sup>1</sup>, Eni Irfiani<sup>2</sup>

<sup>1</sup>Program Studi Sistem Informasi Akuntansi, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika Jakarta

<sup>2</sup>Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika Jakarta

<sup>1</sup>[fintri.fni@bsi.ac.id](mailto:fintri.fni@bsi.ac.id)

<sup>2</sup>[eni.enf@bsi.ac.id](mailto:eni.enf@bsi.ac.id)

**Abstrak** - Menjaga stok persediaan barang agar tidak ada barang yang kosong termasuk salah satu cara untuk menjaga kepuasan pelanggan. Untuk memenuhi hal tersebut penjual harus dapat menganalisa mana data barang yang laku dan mana yang kurang laku dari data laporan penjualan barang, hal ini tidaklah mudah apabila toko tersebut merupakan toko retail yang memiliki ratusan bahkan ribuan data penjualan setiap bulannya. Permasalahan tersebut bisa di selesaikan dengan menggunakan salah satu teknik dalam data mining yaitu algoritma K-Means Clustering. Penelitian ini di maksudkan untuk membantu Genta Corp yang merupakan toko retail di kota Bogor yang menjual peralatan outdoor, untuk membuat pengelompokan data penjualannya agar dapat memaksimalkan manajemen stoknya. Variable yang digunakan adalah kode barang, data barang masuk, data barang keluar dan stok barang. Data diolah dengan perhitungan manual menggunakan algoritma K-Means dan menggunakan Software Rapid Miner sehingga didapatkan hasil akhir berupa tiga cluster dimana terdapat 2 jenis barang paling laris, 8 jenis barang yang cukup laris dan 18 jenis barang yang kurang laris. Hasil ini bisa dimanfaatkan oleh manajemen Genta Corp untuk peningkatan manajemen stok dan strategi penjualannya.

**Kata-kata kunci:** *Clustering, Data Mining, Manajemen Stok, Penjualan, Peralatan Outdoor.*

**Abstract** - Maintain stock of goods so that there are no empty items, including one way to maintain customer satisfaction. To fulfill this, the seller must be able to analyze which data items are selling and which are not selling from the sales report data, this is not easy if the store is a retail store that has hundreds or even thousands of sales data every month. These problems can be solved by using one of the techniques in data mining namely algorithm K-Means Clustering. This research is intended to help Genta Corp which is a retail store in the city of Bogor that sells outdoor

equipment, to make a grouping of sales data in order to maximize stock management. Variables used are item code, item data, item data and item stock. The data is processed by manual calculation using the K-Means algorithm and using the Rapid Miner Software so that the final result is in the form of three clusters where there are 2 types of best-selling goods, 8 types of goods which are quite in demand and 18 types of goods that are not in demand. This result can be utilized by the management of Genta Corp to improve stock management and sales strategies.

**Keywords:** *Clustering, Data Mining, Stock Management, Sales, Outdoor Equipment.*

### I. PENDAHULUAN

Kegiatan mendaki gunung atau berkemah merupakan olahraga atau hobi yang banyak diminati saat ini, dimana kita bisa melakukan olahraga sekaligus refreshing. Akan tetapi tidak sedikit kasus kematian yang sering kita dengar dari aktifitas tersebut, hal ini salah satunya dikarenakan kurangnya perlengkapan standar keamanan yang dimiliki para pecinta kegiatan outdoor. Sehingga saat ini banyak bermunculan toko yang menjual dan atau menyewakan peralatan outdoor. Dalam memenuhi kebutuhan para pecinta alam toko Genta Corp hadir di kota Bogor, menjual beragam kebutuhan outdoor, toko ini memang masih kecil item yang dijual saat ini ada tiga puluh, tetapi tidak menutup kemungkinan akan ada penambahan item seiring dengan peningkatan penjualannya.

Data mining merupakan penggalian data yang tersembunyi dari database. Proses Clustering merupakan suatu proses pengelompokan berdasarkan atas prinsip kesamaan kelas serta mengurangi kesamaan antar kelas. Berbagai algoritma dalam clustering telah dikembangkan untuk menghasilkan kinerja yang baik. Keakuratan

perkiraan7 penjualan memiliki dampak yang besar pada penjualan. Peramalan penjualan menjadi factor penting pada manajemen dimulai dari pengecer sampai pada distributor, produsen dan supplier. Hasil perkiraan penjualan menggunakan metode k-means yang tepat dan akurat dapat menjadi jembatan antara banyak penawaran dan permintaan sehingga mampu mengurangi biaya dan mempertahankan jumlah stok barang [1][2][3].

Kemampuan untuk memperkirakan volume penjualan secara akurat pada setiap item yang dijual di toko sangat penting guna kelangsungan hidup pada bisnis. Beberapa factor penting seperti penetapan harga, alokasi serta pengelolaan inventaris secara langsung terkait dengan perkiraan penjualan. Agar tidak terjadi kesalahan dalam perkiraan penjualan yang berdampak pada hilangnya sejumlah uang, untuk itu diperlukan suatu peramalan yang mendukung pengambilan keputusan. Salah satu metode yang digunakan untuk mengurangi permasalahan adalah clustering. Metode clustering bermanfaat untuk menemukan jumlah cluster yang tepat pada data penjualan [4].

Untuk menjaga kepuasan pelanggan toko Genta Corp berusaha untuk selalu menyediakan peralatan yang dibutuhkan pelanggannya, tetapi pengelolaan data stok barang dan penjualan masih dilakukan secara manual dan belum ada analisa terhadap data penjualan secara menyeluruh sehingga terkadang ada barang yang menumpuk karena kurang laku dan ada barang yang kosong [5]. Sehingga diperlukan sistem yang dapat membantu toko dalam manajemen stok

Dalam melakukan perencanaan stok diperlukan histori data penjualan, apabila data penjualan besar maka di perlukan teknik data mining [6]. Dimana pada penelitian ini di gunakan algorithm K-Means untuk mengelompokkan data berdasarkan data barang yang paling laris, cukup laris dan kurang laris.

Metode K-Means telah banyak di terapkan dalam berbagai kasus selain kasus penjualan, missal pengelompokan jumlah penumpang bus trans Jogja berdasarkan jalur bus dan shelter [7], pengelompokan pelanggan PDAM [8], dan pengelompokan untuk menentukan gizi balita [9].

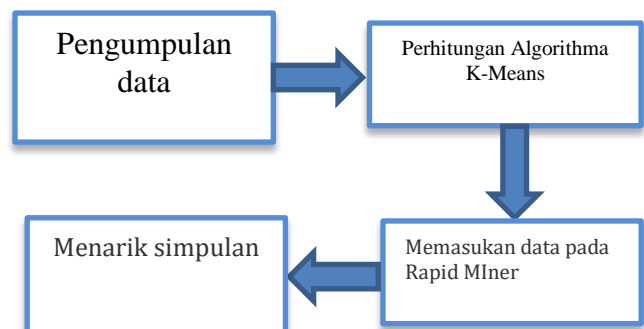
## II. METODE

### A. Tahapan Penelitian

Penelitian ini menggunakan data mining. Di era industri 4.0 ini dimana perkembangan data menjadi bigdata maka di butuhkan metode untuk mengolah bigdata agar bermanfaat bagi kehidupan kita, salah satu metode nya adalah data mining. Dimana di dalam data mining merupakan cara untuk mengotomatisasi proses

guna menemukan pola dari kumpulan data yang berskala besar [10]. Data mining digunakan untuk mengolah bigdata dalam basis data sehingga menghasilkan informasi baru yang berguna bagi strategi bisnis. Secara garis besar data mining dibagi menjadi dua kategori utama yaitu Deskriptive Mining dan Predictive Mining dimana pada descriptive mining digunakan untuk menentukan karakteristik data sedangkan predictive mining digunakan untuk menemukan pola data[11]. dilakukan dengan tahapan-tahapan yang mengacu pada tahapan metode K-Means, K-Means merupakan salah satu metode dalam data mining. Dimana pada metode ini dilakukan pengelompokan data (*clustering*) dengan sistem partisi dan pemodelannya tanpa supervisi [12]. Dalam metode ini pengelompokan data dilakukan menjadi beberapa kelompok, dan tiap kelompok memiliki kesamaan karakteristik dan memiliki karakteristik dengan kelompok lainnya [13].

Algorithma K-Means melakukan clustering berbasis titik (*centroid*) dengan menentukan tiga parameter yaitu jumlah cluster, inialisasi cluster dan jarak sistem [14]. yang disesuaikan dengan objek penelitian yang ada baik secara manual maupun dengan bantuan software Rapid Miner, berikut adalah tahapannya (Gambar 1).



**Gambar 1. Tahapan penelitian**

Langkah yang dilakukan untuk membentuk clustering pada metode K-Means adalah:

1. Menentukan banyaknya cluster (k) untuk jumlah cluster dari dataset yang ada.
2. Menentukan k sebagai Centroid, biasanya dilakukan secara acak (random).
3. Hitung jarak data dengan centroid menggunakan rumus jarak menggunakan rumus Euclidean (persamaan 1)

$$\text{Distance: } d(x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2} \quad (1)$$

dimana d merupakan titik dokumen,  $x_i$  merupakan data kriteria dan  $\mu_j$  merupakan centroid pada cluster ke-j.

- Kelompokkan data berdasarkan kedekatan dengan centroid kemudian perbaharui nilai centroid baru dengan lokasi dari pusat cluster menggunakan persamaan 2:

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (2)$$

Dimana  $\mu_j(t+1)$  merupakan centroid baru pada iterasi ke  $(t+1)$  dan  $N_{sj}$  merupakan banyaknya data pada cluster  $s_j$ .

- Lakukan langkah 2 sampai 4 sampai anggota tiap cluster tidak ada yang berubah.

### B. Objek Penelitian

Objek penelitian adalah Genta Corp, Genta corp adalah sebuah toko yang menjual peralatan outdoor yang berlokasi di kota Bogor. Data diambil dari data persediaan barang selama tiga bulan dengan item data barang sebanyak tiga puluh, data berupa data kuantitatif yang dapat dihitung secara langsung sebagai variable bilangan atau angka. Jenis data merupakan data primer, yang dimaksud dengan data primer adalah data didapat langsung dari Genta Corp dengan metode wawancara dengan karyawan dan observasi. Populasi data yang digunakan adalah data stok barang selama tiga bulan dari bulan April, Mei dan Juni 2019. Dengan data master barang sebanyak tiga puluh dan jumlah transaksi 140 transaksi. Perhitungan sampel menggunakan rumus slovin [15] dengan tingkat error sebesar 5%, maka didapat perhitungan sebagaimana persamaan 3:

$$n = N / (1 + (N \times e^2)) \quad (3)$$

$$n = 140 / (1 + (140 \times 0.5^2))$$

$$n = 140 / (1 + (140 \times 0.05))$$

$$n = 140 / 8$$

$$n = 17,5 \text{ atau dibulatkan menjadi } 18$$

## III. HASIL DAN PEMBAHASAN

### A. Pengumpulan Data

Dari data master barang sebanyak 30 jenis barang maka akan diolah dengan menggunakan algorithm K-Means (Tabel 1).

### B. Perhitungan Algorithm K\_Means

Langkah pertama dilakukan dengan menentukan menentukan banyaknya cluster dan nilai centroid awal, jumlah cluster ditentukan berdasarkan variable pengelompokan data penjualan laris, cukup laris dan kurang laris. langkah berikutnya menentukan centroid awal (k), dapat diambil data secara random atau acak (Tabel 2).

TABEL I  
DATA STOK BARANG GENTA CORP

Kode Barang	IN	OUT	STOK
B001	20	10	10
B002	7	2	5
B003	6	1	5
B004	20	5	15
B005	5	2	3
B006	15	10	5
B007	25	10	15
B008	10	5	5
B009	8	3	5
B010	10	7	3
B011	15	5	10
B012	15	5	5
B013	25	5	20
B014	25	15	10
B015	7	2	5

TABEL II  
NILAI AWAL CENTROID

Centroid	IN	OUT	STOK
C1	100	50	50
C2	25	15	10
C3	6	1	5

Langkah ke tiga dan empat, dari tiga nilai centroid diatas, lalu kita hitung jarak data dengan centroid dengan rumus *Euclidean Distance* dan langsung kita kelompokkan data berdasarkan kedekatan (similaritas)nya, sehingga didapat data sebagai berikut (Tabel 3).

Dari langkah ke tiga dan keempat kita hitung rata-rata centroidnya sehingga mendapat nilai centroid baru yaitu untuk C1 65, 27.5, 37.5, C2 23.3, 8.9, 14.4, C3 10.6, 4.4, 6.8 kemudian kita ulangi prosesnya mulai dari langkah kedua, sehingga didapat hasil sebagai berikut (Tabel 4).

Setelah iterasi kedua maka dihitung kembali rata-rata centroid sehingga di peroleh nilai centroid baru C1 sebesar 65, 27.5, 37.5, C2 sebesar 24.4, 8.8, 15.6, C3 sebesar 10.9, 4.7, 6.7 dan dikarenakan pada iterasi kedua masih ada anggota cluster yang berubah maka dilanjutkan ke iterasi ketiga. Dengan hasil sebagai berikut (Tabel 5).

Dari Tabel 5 nilai anggota cluster sudah tidak ada yang berubah sehingga kita sudah sampai tahap terakhir. Jika dilakukan perhitungan prosentase untuk centroid C1 sebesar 64.7% C2 24.3% dan C3 11.1% sehingga dapat di simpulkan bahwa item data yang

berada di cluster C1 merupakan barang yang paling laris, yang berada di C2 adalah barang yang cukup laris dan yang berada di C3 adalah barang yang kurang laris.

TABEL III  
JARAK DAN SIMILARITAS CENTROID PADA ITERASI I

Kode Barang	JC1	JC2	JC3	C1	C2	C3
B001	97,979	7,071	17,378		*	
B002	113,921	22,759	1,414			*
B003	115,161	24,125	0			*
B004	98,234	12,247	17,663		*	
B005	116,353	24,859	2,449			*
B006	104,163	12,247	12,728		*	
B007	91,924	7,071	23,281		*	
B008	110,227	18,708	5,657			*
B009	112,685	21,401	2,828			*
B010	110,263	18,385	7,483			*
B011	104,163	14,142	11,045			*
B012	106,184	15	9,849			*
B013	92,466	14,142	24,536		*	
B014	91,924	0	24,125		*	
B015	113,921	22,759	1,414			*

TABEL IV  
JARAK DAN SIMILARITAS CENTROID PADA ITERASI II

Kode Barang	JC1	JC2	JC3	C1	C2	C3
B001	55,565	5,666	11,394		*	
B002	71,207	20,086	4,685			*
B003	72,384	21,257	5,996			*
B004	55,113	5,152	12,479		*	
B005	73,759	22,683	7,183			*
B006	62,149	12,644	7,345			*
B007	49,117	2,079	17,485		*	
B008	67,731	16,796	1,988			*
B009	70,039	18,947	3,457			*
B010	68,084	17,673	4,647			*
B011	61,339	10,214	5,462			*
B012	63,738	13,182	4,741			*
B013	49,117	6,983	19,535		*	
B014	50,125	7,738	18,161		*	
B015	71,207	20,086	4,685			*

TABEL V  
JARAK DAN SIMILARITAS CENTROID PADA ITERASI III

Kode Barang	JC1	JC2	JC3	C1	C2	C3
B001	55,565	7,235	11,071		*	
B002	71,207	21,456	5,001			*
B003	72,384	22,596	6,334			*
B004	55,113	5,796	12,352		*	
B005	73,759	24,090	7,432			*
B006	62,149	14,225	6,937			*
B007	49,117	1,531	17,234		*	
B008	67,731	18,265	1,918			*
B009	70,039	20,349	3,728			*
B010	68,084	19,212	4,436			*
B011	61,339	11,558	5,303			*
B012	63,738	14,658	4,449			*
B013	49,117	5,796	19,417		*	
B014	50,125	8,432	17,805		*	
B015	71,207	21,456	5,001			*

### C. Memasukan Data pada Rapid Miner

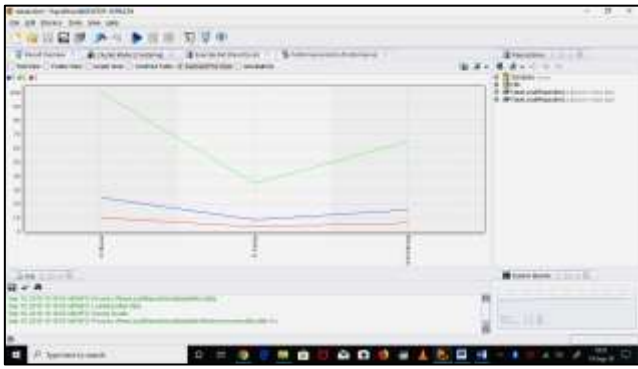
Setelah melakukan perhitungan manual dengan algoritma K-Means, kemudian di compare dengan menggunakan aplikasi Rapid Miner maka di dapat hasil yang kurang lebih sama, maka dapat di simpulkan bahwa perhitungan yang dilakukan sudah sesuai. Berikut adalah hasil pengolahan data penjualan dengan menggunakan rapid miner (Gambar 2).



Gambar 2. Hasil perhitungan dengan Rapid Miner

Adapun jika digambarkan dengan grafik hasilnya menjadi sebagai berikut (Gambar 3).





Gambar 3. Grafik K-Means dengan Rapid Miner

#### IV. PENUTUP

Penerapan metode K-Means dalam pengelompokan data penjualan pada Toko Genta Corp dapat menghasilkan rekomendasi barang yang laris, Kurang laris dan cukup laris. Sehingga data dijadikan rujukan bagi manajemen untuk mengatur stok barang agar toko tidak mengecewakan pelanggan karena barang yang ingin di beli tidak tersedia. Untuk pengembangan lebih lanjut dan agar pihak toko mudah dalam penggunaan metode ini maka perlu dibuat sebuah aplikasi dengan acuan metode K-Means. Apabila dirasa perlu dapat dikembangkan menjadi Expert System.

#### DAFTAR PUSTAKA

- [1] M. Sanwlani, "FORECASTING SALES THROUGH TIME SERIES CLUSTERING," vol. 3, no. 1, pp. 39–56, 2013.
- [2] V. Shrivastava, P. Arya, and M. T. S. Systems, "International Journal of Computing , Communications and Networking Available Online at <http://warse.org/pdfs/ijccn04122012.pdf> A Study of Various Clustering Algorithms on Retail Sales Data," 2012.
- [3] A. Benet-zepf, J. A. Marin-garcia, and I. Küster, "Clustering the mediators between the sales control systems and the sales performance using the AMO model : A narrative systematic literature review," vol. 14, no. 2, pp. 387–408, 2018.
- [4] M. Kumar and N. R. Patel, "Using clustering to improve sales forecasts in retail merchandising," pp. 33–34, 2010.
- [5] B. M. Metisen and H. L. Sari, "ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA," vol. 11, no. 2, pp. 110–118, 2015.
- [6] E. Muningsih and S. Kiswati, "Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang," vol. 3, no. 1, 2015.
- [7] L. Zahrotun, "ANALISIS PENGELOMPOKAN JUMLAH PENUMPANG BUS TRANS JOGJA MENGGUNAKAN METODE CLUSTERING K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING ( AHC )," vol. 9, no. 1, pp. 1039–1047, 2015.
- [8] S. T. Siska, "ANALISA DAN PENERAPAN DATA MINING UNTUK MENENTUKAN KUBIKASI AIR TERJUAL BERDASARKAN PENGELOMPOKAN PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," vol. 9, no. 1, pp. 86–93, 2016.
- [9] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," vol. 6, no. 4, pp. 161–168, 2018.
- [10] M. H. Siregar, "KLAUSTERISASI PENJUALAN ALAT-ALAT BANGUNAN MENGGUNAKAN METODE K-MEANS (STUDI KASUS DI TOKO ADI BANGUNAN)," vol. 1, no. 2, pp. 83–91, 2018.
- [11] Y. Darmi and A. Setiawan, "PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK," vol. 12, no. 2, pp. 148–157, 2016.
- [12] S. P. Tamba, F. T. Kesuma, and Feryanto, "PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING," vol. 2, no. 2, 2019.
- [13] Mardalius, "PENGELOMPOKAN DATA PENJUALAN AKSESORIS MENGGUNAKAN ALGORITMA K-MEANS," vol. IV, no. 2, pp. 401–411, 2018.
- [14] Susliansyah, H. Sumarno, H. Priyono, and N. Hikmah, "Pengelompokan Data Pembelian Tinta Dengan Menggunakan Metode K-means," vol. 3, no. September, pp. 381–392, 2019.
- [15] Sugiyono, "Metode Penelitian Kuantitatif, Kualitatif, dan R&D," Bandung: CV Alfabeta, 2017.

